

Gerenciamento de Conteúdo informação não estruturada a salvo de tragédias

Francisco Camargo

AGENDA

1. Enterprise Content Manager
2. Recuperando conteúdos não estruturados, mas essenciais, em caso de desastre.
3. Evitando o vazamento de conteúdo confidencial.

Conteúdo organizacional



O Conteúdo deve ser gerenciado como um recurso estratégico da empresa.

Conteúdo organizacional

Qualquer “conteúdo” a ser arquivado, e que seja consultado por alguém, deve ser considerado Conteúdo Organizacional!

Manuais, Especificações, Procedimentos, Propostas, Memorandos, Cartas, Arquivos CAD, Contratos, Fax, e-mail, Relatórios Financeiros, Relatórios de Consultoria, Áudio, Vídeo, DLLs, Objetos COM+, Programas, Balanços, Documentos Legais, Registros de Pessoal ...

1

Gerenciamento de Conteúdo

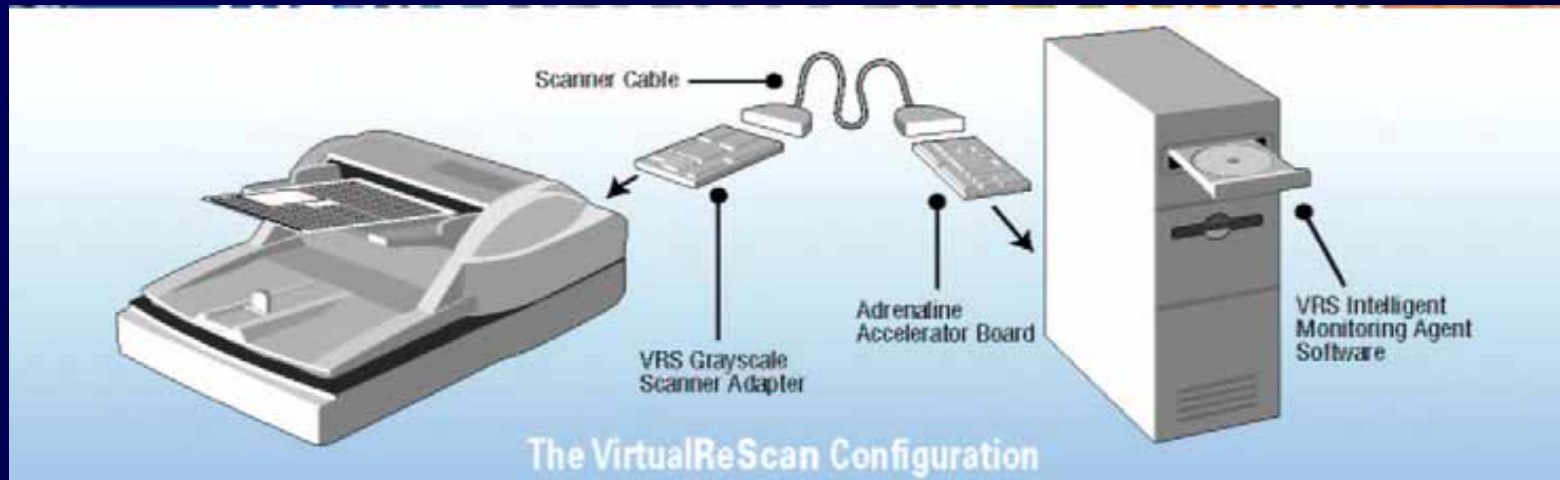
GED



Gerenciamento Eletrônico de Documentos

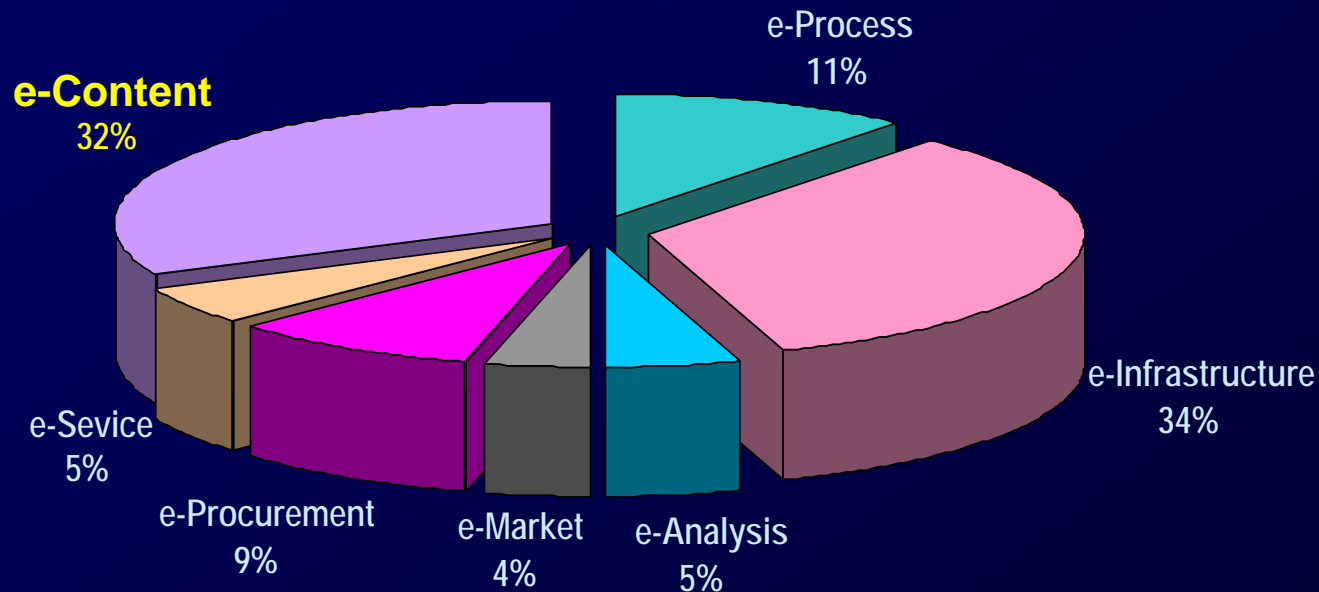
O início do Enterprise Content Management.

Os documentos em papel, passam por um **SCANNER**, que captura a **imagem** do documento e a seguir por um **OCR** (optical character recognition) que transforma o documento em **texto**, pronto para ser armazenado digitalmente.



Estatística: Mercado B2B

Ferramentas voltadas para B2B
Tamanho do mercado = US\$ 6 Bilhões



Os segmentos com crescimento mais acelerado são
e-Infrastructure, e-Markets e e-Content

Arquivamento

- Só no Repositório
- O Repositório não é acessível a não ser através do Gerenciador de Conteúdo
- Os documentos só são acessíveis pelo Gerenciador de Conteúdo
- Qualquer documento do tipo *.*

Autenticação de Usuário

- Um dos pilares da segurança fornecida pelo Gerenciador de Conteúdo
- Identifica a que grupos de segurança o usuário tem ou pertence
- Que papel exerce no sistema, leitor, contribuidor, aprovador ou administrador
- Leitores anônimos (documentos públicos)

Fonte Única

- Para que manter **várias cópias**, nem sempre atualizadas de um documento na Organização?
- Muito mais inteligente seria manter um **documento único**, atualizado, em um único lugar e divulgar só o seu **endereço**.
- Cada documento do **Repositório** passa a ter um **endereço fixo**, ou na linguagem da Internet, uma **URL Persistente**.

Metadados

Metatag: conceito emprestado da WEB

- Não visíveis ao leitor comum, informações relevantes sobre o documento, permitindo classificá-lo, armazená-lo, e encontrá-lo quando for necessário.
- Vídeo clipe (streaming media), em que os campos seriam o Nome do Vídeo, do Diretor, o Assunto, a Duração, os Artistas, etc..
- Imagens precisam de **Metadados** bem especificados, para serem facilmente localizadas
- Os documentos com texto: o mecanismo de busca pode localizá-los por palavras chave que fazem parte do texto ou pode efetuar uma indexação **FULL TEXT**



Disponibilidade

IDC research indicates that Fortune 500 companies will lose \$31.5 billion by 2003 due to the inability to find information quickly and efficiently.

Source: IDC, as reported by *Computerworld*. July 18, 2001.

500 Maiores Empresas do mundo perderão 31,5 bilhões de dólares em 2003 pela dificuldade de encontrarem a informação certa, rápida e eficientemente.

Busca (Full text retrieval)

- De nada adianta ter todos os documentos arquivados no repositório, se não se consegue encontrá-los facilmente.
- Para isso os documentos são indexados pelos Metadados, que funcionam como “Palavras-Chave” .
- O Banco de Dados é varrido e rapidamente os **endereços dos documentos** que contenham essa palavra-chave são encontrados.
- Nem todos os documentos encontrados serão mostrados ao usuário, somente àqueles a que ele tem direito de acesso, em razão do Sistema de Segurança.
- Para os documentos constituídos principalmente de texto, é usual a indexação pela totalidade documento.

Subscrição

- Com todos os usuários identificados, é possível a eles se inscreverem ou assinarem, “Boletins” de determinadas áreas.
- Quando uma nova informação é publicada em uma área em que o usuário tenha interesse e na qual ele se inscreveu, o sistema pode mandar um e-mail, com o endereço do novo documento, com os seguintes dizeres:
“Você subscreveu o Grupo de Normas. O seguinte documento foi acrescentado a este grupo nesta data:
<http://mercury/normas/N430.html> ”
- Isto torna muito cômodo a confecção de **Newsletters** e **Revistas Eletrônicas**.

Publicação automática

Separando Forma do Conteúdo

Forma e *Conteúdo*

Forma = Estética e Navegação

Conteúdo = Informação

O mesmo documento pode ser publicado em diferentes sites, diferentes intranets, com diferentes formas, mas com exatamente o mesmo conteúdo.

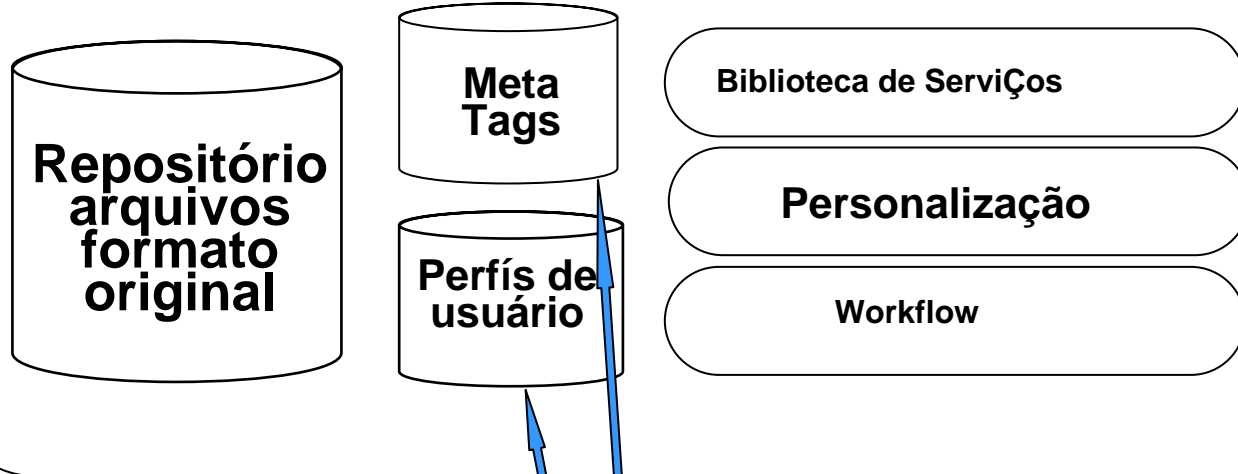
Personalização

- Por causa da segurança, os usuários e as informações foram identificados e devidamente classificados.
- Uma vez identificados, é possível personalizar as áreas (as informações) a que o usuário vai ter acesso.
- É possível criar sites que sejam específicos e totalmente personalizados para cada usuário.
- Quando o usuário da Internet / Intranet faz o Log In, automaticamente o sistema o identifica e personaliza a sua página de entrada e outras.

Arquitetura

FILE
SYSTEM

Gerenciador de Conteúdo Motor de Gerenciamento para Dados não estruturados



Publicador de conteúdo

Controla o look & feel

Banco de
dados



Gerenciador de Conteúdo

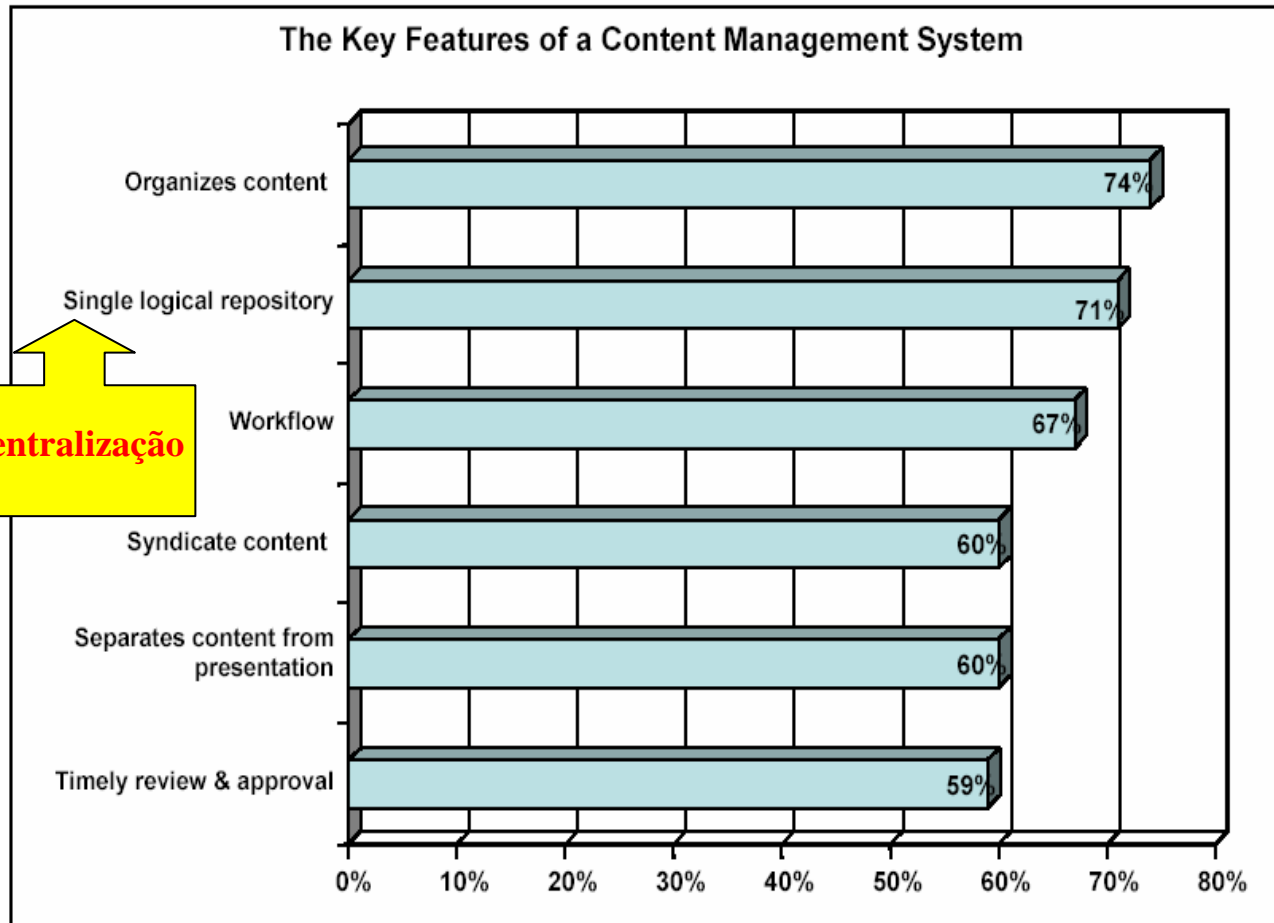
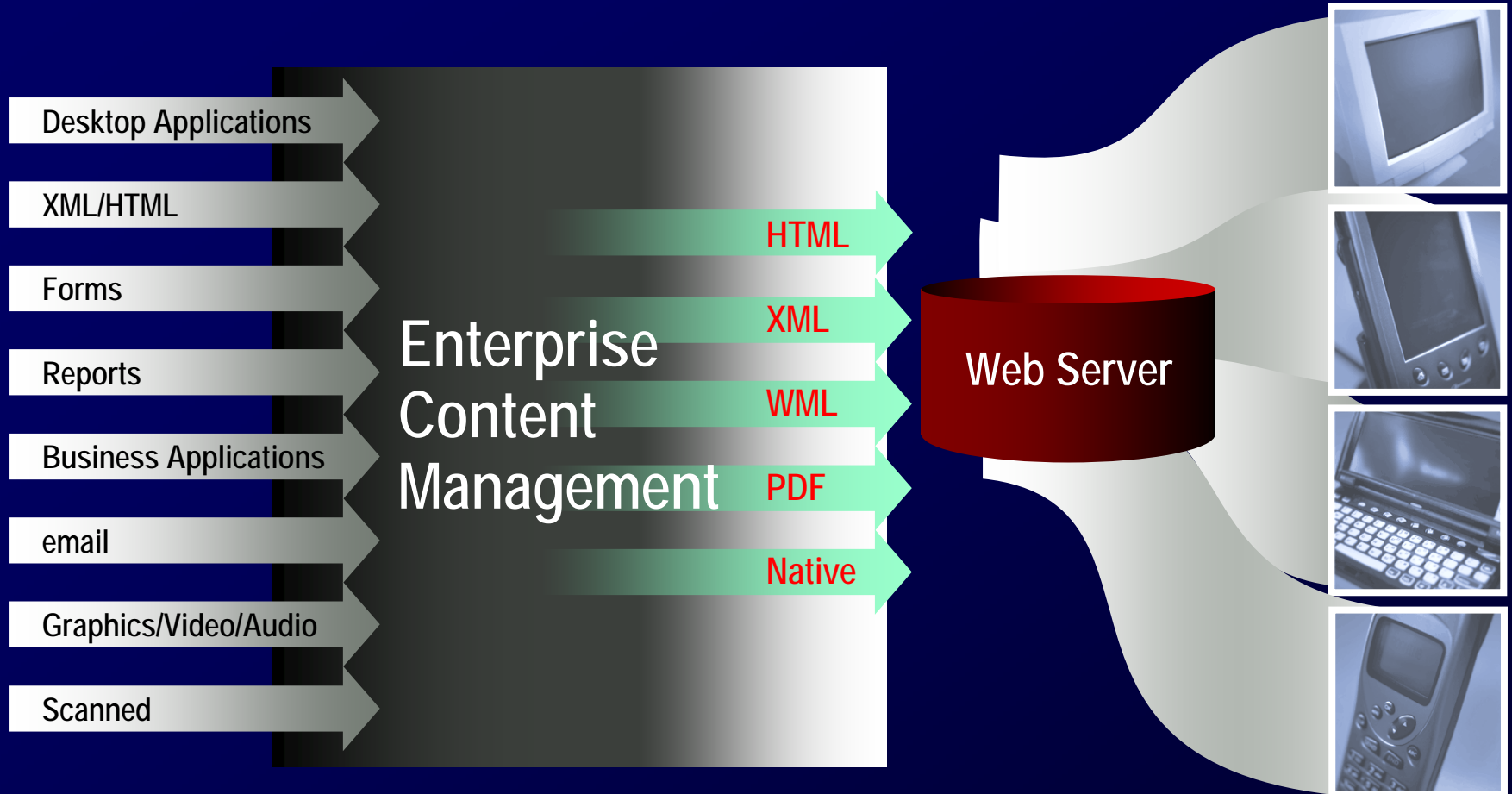


Illustration 3. Our respondents identified a number of key features for a content management system. (Multiple responses were allowed.)

Enterprise Content Management



Conteúdo de diversas fontes de recursos

Conversão de diversos formatos em formatos publicáveis na web

Segurança e personalização

2

Recuperando conteúdos não estruturados,
mas essenciais, em caso de desastre.



A Secure, Resilient Enterprise needs Content Management

Gartner Group

Uma empresa, segura e resistente,
precisa de um Gerenciador de Conteúdo



Informação não Estruturada

Não é informação inútil ou sem importância:

Memorandos, Cartas, Propostas,
Contratos, Website ...

Gartner estimates that unstructured information doubles every three months.

Source: Gartner, as reported by *Transform*. May 2001.

Informação não Estruturada

O desafio do
papel:

Muita informação
não estruturada

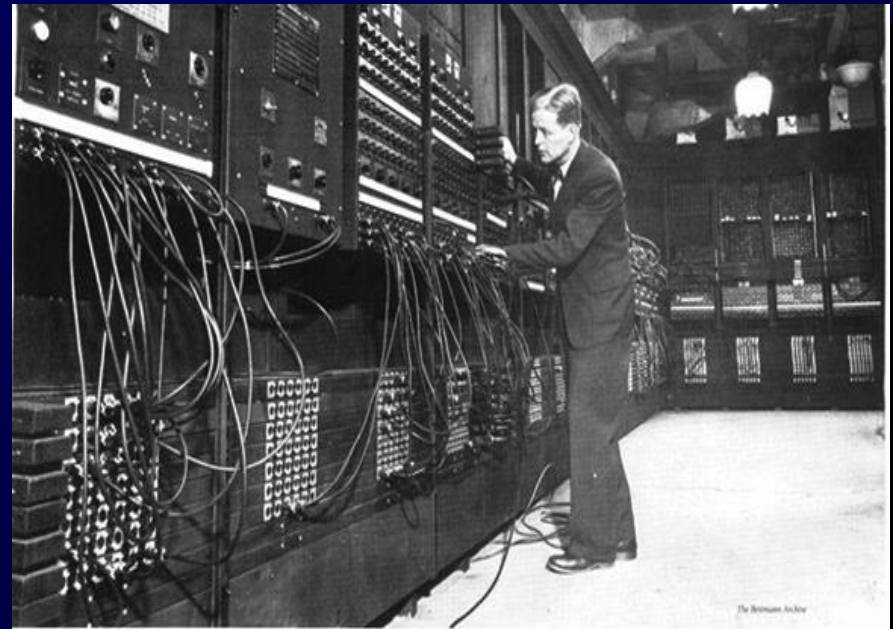
é arquivada na forma de
documentos em papel,
sujeitos a incêndio,
perda, roubo, acesso
não controlado.



Informação não Estruturada

O desafio do
drive C:

Muita informação
não estruturada,
é armazenada no drive C,
sem controle, sem senha,
sem backup



Informação não Estruturada

When planning for disaster recovery, enterprises seldom include unstructured content, such as paper and electronic documents, Web content, and Intranet content.

Debra Logan
Research Analyst Gartner

Quando desenvolvem um Plano de Continuidade de Negócios (Recuperação de Desastres), as empresas, raramente, incluem o Conteúdo não-estruturado, tais como documentos eletrônicos, documentos em papel, conteúdo dos sites web e intranet.

“O acaso faz da vida um jogo...”

Lucius Sêneca, 1 DC

.... de cada cinco
empresas que sofreram
interrupção nas suas
operações por uma
semana, duas fecharam
as portas em menos de
três anos.

DRI - Disaster Recovery Institute



Gerenciador de Conteúdo

- O Gerenciador de Conteúdo centraliza as informações em um repositório, que pode ser replicado em mais de um site.
- Centralização, Organização, Categorização

A informação não estruturada está a salvo em um Gerenciador de Conteúdo !

3

Evitando o vazamento de conteúdo confidencial.

Vazamento de Informação

Decreto nº 4.553, de 27 de dezembro de 2002, dispõe sobre a salvaguarda de dados, informações, documentos e materiais sigilosos de interesse da segurança da sociedade e do Estado, no âmbito da Administração Pública Federal, e dá outras providências, trata também sobre a classificação segundo o grau de sigilo:

Art. 5º Os dados ou informações sigilosos serão classificados em ultra-secretos, secretos, confidenciais e reservados, em razão do seu teor ou dos elementos intrínsecos.

O Fator Humano

- Classificação dos documentos quanto ao grau de **Confidencialidade**
- Classificação dos funcionários quanto ao seu nível de **Confiança**.

O Fator Humano

o Ser Humano é
elo mais fraco da
Segurança da Informação

“The Art of Deception: Controlling the Human
Element of Security”

Kevin D. Mitnick *

* Famoso Hacker, esteve preso por 5 anos nos EUA.

Perda por tipo de ataque

TYPE OF ATTACK	AVERAGE LOSS	
Unauthorized Insider Access	\$	1,363,915
Theft of Proprietary Info	\$	1,307,146
Financial Fraud	\$	656,927
Telecom Fraud	\$	595,766
Sabotage of data or networks	\$	164,817
Spoofing	\$	128,000
System penetration by outside	\$	110,944
Telecom Eavesdropping	\$	96,833
Denial of Service	\$	77,417
Virus	\$	65,997
Active Wiretapping	\$	49,000
Insider Abuse of Net Access	\$	38,744
Laptop Theft	\$	35,348
AVERAGE LOSS	\$	215,753

2001 - Fonte Verisign

O Fator Humano

- Classificação dos documentos quanto ao grau de **Confidencialidade**
- Classificação dos funcionários quanto ao seu nível de **Confiança**.

O Fator Humano

A política de segurança deve indicar os procedimentos para **classificar os conteúdos e funcionários**, permitindo que os funcionários acessem só o que tenham direito.



Vazamento de Informação

Toda Informação é confidencial !

Matriz CRUD

A Informação só pode ser

Created – Criada

Read – Lida

Updated – Atualizada

Deleted – Apagada

E.F. Cood

Cada fase acima é um problema de segurança.

Gerenciador de Conteúdo

Leitura de Documentos

Autenticação de Usuário

Criação de Documentos

Workflow

Deleção de Documentos

Controle de versões



Gerenciador de Conteúdo

- **Confidencialidade** - os dados devem ser lidos somente por pessoal autorizado a fazê-lo.
- **Integridade** - os dados não devem sofrer modificações não autorizadas.
- **Disponibilidade** - os dados estarão acessíveis quando os usuários necessitarem.
- **Rastreabilidade** – Trilha de Auditoria, para identificar quem fez o que.



Gerenciador de Conteúdo

As pessoas só
respeitarão as
diretivas e políticas
que lhes forem
impostas, de
preferência por meio
de sistemas
automáticos

De nada adiantam:
Projetos de
Conscientização,
Campanhas,
Circulares,
Manuais, etc.



Gerenciador de Conteúdo

Uma Política de Segurança da Informação pode ser mais facilmente implementada e respeitada através de uma ferramenta de Gerenciamento de Conteúdo.

Uma mensagem de otimismo:



Você, cuidadosamente, identificou todas as vulnerabilidades e configurou a arquitetura da sua rede para ficar imune a elas.

Você implantou Antivirus, Firewalls e um IDS.

Todas as suas diretivas de rede e procedimentos estão em ordem.

Você sabe o que está fazendo, pois você foi treinado para fazê-lo.

Será que nada pode dar errado?

Pense outra vez !



ERA DA INTERNET



"The wonder of the Web is that the customer knows about IT problems the same time you do. There's no camouflage."

"O que é maravilhoso na era da Internet é que os seus clientes ficam sabendo dos problemas de informática da sua empresa, quase ao mesmo tempo que voce.

Não há camuflagem."

*Senior Vice President
electronic brokerage technology*

THE ~~END~~ BEGIN



OBRIGADO

Camargo

fcamargo@clm.com.br

www.clm.com.br

APENDICE



Gerenciador de Conteúdo Universal

breve descrição

Universal Content Management



APPLICATION

Web Sites

Contracts HR Customer Support

Intranets

E-Learning Portal

Extranets

ERP / CRM Integration

Portals

Collaboration File Server Replacement

INTEGRATION

Web Services

EJB / J2EE

JSP Tag Library

Com

Scripting

Command Line

ICE

HTTP

APPLICATION SERVICES

WCM

DOCUMENT
MANAGEMENT

COLLABORATION

FILE SERVICE

DAM / RM

CORE SERVICES

LIBRARY
SERVICES

CONVER-
SION

SECURITY

PERSONAL-
IZATION

WORK-
FLOW

RECORD
MGMT

INPUT

ADMIN-
ISTRATION

Web content management requirements



- **Site design**
- **Publishing**
- **Deployment**
- **Templates**
- **In-context edits**
- **Site migration**

Document management requirements



- **Web-based**
- **Workflows**
- **Compound docs**
- **XML authoring**
- **Doc merge**
- **Imaging**

Collaboration requirements

COLLABORATION

- **Ad hoc workflow and security**
- **Project portal**
- **Integrate with calendar, PM, live collaboration**

File service requirements

- **Scale & performance**
- **Folders**
- **Mapped drives**
- **Desktop app integrations**
- **Explorer integration**

A diagram illustrating the requirements for a file service. On the left, a white rectangular box with a dashed orange border contains a bulleted list of requirements: Scale & performance, Folders, Mapped drives, Desktop app integrations, and Explorer integration. A horizontal dashed orange line connects the right side of this box to a solid light green rectangular box on the right. Inside the green box, the text 'FILE SERVICE' is written in a bold, green, sans-serif font.

FILE SERVICE

Digital asset / Records management requirements



- **Audio indexing**
- **Video indexing**
- **Streaming video**
- **Thumbnails**
- **Compression**
- **Image conversion**

DAM / RM

Library services requirements

- **Check-in / check-out**
- **Revision management**
- **Search and retrieval**
- **Metadata management**

LIBRARY
SERVICES

Conversion services requirements



- From native formats and XML
- To HTML
- To XML
- To PDF
- To Wireless

CONVER-
SION



Security requirements

SECURITY

- Integrate with existing directory services – LDAP, AD, other
- Role-based
- Rule-based
- Doc Level
- Usage analytics & reporting

Personalization requirements

- **Security-Defined**
- **User-Defined**
- **Administrator-Defined**
- **Rules-based**

PERSONAL-
IZATION

Workflow requirements

- GUI
- Integrate with external BP
- Variable Driven

A light green rectangular box containing the text 'WORKFLOW' in a bold, black, sans-serif font, arranged in two lines: 'WORK-' on the top line and 'FLOW' on the bottom line.

WORK-
FLOW

Records management requirements



- Archival and Retrieval
- Disposition
- Certification
- Subsystem integration

RECORD
MGMT

.br

Input requirements

- **Desktop**
- **Browser**
- **WebDAV**
- **Batch**
- **Report Parser**
- **Scanning**
- **Email**
- **Categorizer**

A light green rectangular box with the word 'INPUT' in bold, black, uppercase letters. A dashed orange line connects the right side of the list box to the left side of this box.

INPUT

Administration requirements



- **Web-based**
- **Multi-server mgmt**
- **Clustering**
- **Enterprise search**
- **Analytics**
- **Replication**
- **Archiving**

**ADMIN-
ISTRATION**

Open enterprise architecture



SCALABLE ARCHITECTURE

Presentation Layer	Portals and Application Servers					
Content and Code Layer	HTML	JSP	ASP	XML	WML	Others
HTTP Server	IIS, NetScape, Apache					
Programming Layer	COM	J2EE	WebDAV	JAVA	SOAP	
Content Management				RDBMS	Full Text	Storage
				JDBC	FT API	File System
Java Virtual Machine	Java Virtual Machine					
Operating System	HP IBM Sun UNIX, W2K, NT, Linux					



Quem é a CLM ?

Distribuidor de Soluções
nas áreas de Gerenciamento de Risco
Operacional, Segurança da Informação e
Webmetricas.

Av. Ibirapuera, 2120 – 9. andar

Tel. 11-21***CLMCLM***

(11-2125-6256)

www.clm.com.br